

Species sampling has a major impact on phylogenetic inference

G. Lecointre, H. Philippe, H.L. Van Lê, H. Le Guyader. 1993. *Molecular Phylogenetics and Evolution*. 2 : 205-224.

Representative properties of gnathostome species of a rich 28S rRNA data base were studied through the analysis of the fluctuations they provoked in bootstrap proportions (Bps) of nodes of parsimonious trees. Using original programs which permit BP comparison between different trees, it is empirically demonstrated that 4-to 24-species-trees are highly sensitive to species sampling : the inferences obtained from subsets of 4, 8, 16, or 24 species are not congruent with the whole set of 31 species. Study of trees obtained from exhaustively sampling all combinations of single species taken from each presumed monophyletic group shows precisely the impact of each species on the BP of each node. This procedure also shows that the impact of species changes within a given group on tree Bps is localized to its two or three neighboring nodes. The observation of differing impacts of species emphasizes the importance of sampling several species per presumed monophyletic group. It is also concluded that it is necessary to sample several successive outgroups and that the impact of a species on Bps depends mainly on the sampling context. Before undertaking extensive sequencing, the impact of species should be more often considered, since its effect on Bps is stronger than previously thought.